

## RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical  
Information Center (STIC) no errors detected.

Application Serial Number: 10500, 936  
Source: IFW/b  
Date Processed by STIC: 9/29/06

# *ENTERED*



IFW16

## RAW SEQUENCE LISTING

DATE: 09/29/2006

PATENT APPLICATION: US/10/500,936

TIME: 11:36:44

Input Set : A:\717\_2\_PCTseq.txt

Output Set: N:\CRF4\09292006\J500936.raw

4 <110> APPLICANT: Genencor International, Inc.  
 5 Poulouse, Ayrookaran J.  
 7 <120> TITLE OF INVENTION: Multiply-Substituted Protease Variants  
 10 <130> FILE REFERENCE: GC717-2-PCT  
 C--> 12 <140> CURRENT APPLICATION NUMBER: US/10/500,936  
 C--> 13 <141> CURRENT FILING DATE: 2005-03-25  
 15 <150> PRIOR APPLICATION NUMBER: US 60/350,222  
 16 <151> PRIOR FILING DATE: 2002-01-16  
 18 <160> NUMBER OF SEQ ID NOS: 10  
 20 <170> SOFTWARE: FastSEQ for Windows Version 4.0  
 22 <210> SEQ ID NO: 1  
 23 <211> LENGTH: 1494  
 24 <212> TYPE: DNA  
 25 <213> ORGANISM: Bacillus amyloliquefaciens  
 27 <400> SEQUENCE: 1  
 28 ggtctactaa aatattattc catactatac aattaataca cagaataatc tgtctattgg 60  
 29 ttattctgca aatgaaaaaa aggagaggat aaagagttag aggcaaaaaa gtatggatca 120  
 30 gtttgctggt tgcttttagcg ttaatcttta cgatggcggt cggcagcaca tcctctgccc 180  
 31 aggcggcagg gaaatcaaac ggggaaaaga aatatattgt cgggttttaa cagacaatga 240  
 32 gcacgatgag cgccgctaag aagaaagatg tcatttctga aaaaggcggg aaagtgcaaa 300  
 33 agcaattcaa atatgtagac gcagcttcag ctacattaaa cgaaaaagct gtaaaagaat 360  
 34 tgaaaaaaga cccgagcgtc gcttacgttg aagaagatca cgtagcacat gcgtacgcgc 420  
 35 agtccgtgcc ttacggcgta tcacaaatta aagcccctgc tctgcactct caaggctaca 480  
 36 ctggatcaaa tgttaaagta gcggttatcg acagcgggat cgattcttct catcctgatt 540  
 37 taaaggtagc aggcggagcc agcatggttc cttctgaaac aaatccttct caagacaaca 600  
 38 actctcacgg aactcacgtt gccggcacag ttgcggtctt taataactca atcggtgtat 660  
 39 taggcgttgc gccaaagcga tcaactttac ctgtaaaagt tctcggtgct gacggttccg 720  
 40 gccaatagag ctggatcatt aacggaatcg agtgggcgat cgcaacaat atggacgtta 780  
 41 ttaacatgag cctcggcgga cttctgtggt ctgctgcttt aaaagcggca gttgataaag 840  
 42 ccgttgcatc cggcgctcga gtcgttgccg cagccggtaa cgaaggcact tccggcagct 900  
 43 caagcacagt gggctaccct ggtaaatacc cttctgtcat tgcagtaggc gctgttgaca 960  
 44 gcagcaacca aagagcatct ttctcaagcg taggacctga gcttgatgtc atggcacctg 1020  
 45 gcgtatctat ccaaagcacg cttcctggaa acaaatacgg ggcgtacaac ggtacgtcaa 1080  
 46 tggcatctcc gcacgttgcc ggagcggctg ctttgattct ttctaagcac ccgaactgga 1140  
 47 caaacactca agtccgcagc agtttagaaa acaccactac aaaacttggt gattctttct 1200  
 48 actatggaaa agggctgac aacgtacagg cggcagctca gtaaaacata aaaaaccggc 1260  
 49 cttggccccc cggttttttt atttttcttc ctccgcagt tcaatccgct ccataatcga 1320  
 50 cggatggctc cctctgaaaa ttttaacgag aaacggcggg ttgacccggc tcagtcccgt 1380  
 51 aacggccaag tcctgaaacg tctcaatcgc cgcttcccgg tttccggtca gctcaatgcc 1440  
 52 gtaacggtcg gcggcgtttt cctgataccg ggagacggca ttcgtaatcg gatc 1494  
 54 <210> SEQ ID NO: 2  
 55 <211> LENGTH: 382  
 56 <212> TYPE: PRT

(pg. 6)

## RAW SEQUENCE LISTING

DATE: 09/29/2006

PATENT APPLICATION: US/10/500,936

TIME: 11:36:44

Input Set : A:\717\_2\_PCTseq.txt

Output Set: N:\CRF4\09292006\J500936.raw

```

57 <213> ORGANISM: Bacillus amyloliquefaciens
59 <220> FEATURE:
60 <221> NAME/KEY: VARIANT
61 <222> LOCATION: 163, 164
62 <223> OTHER INFORMATION: Xaa = Pro or Asn
64 <221> NAME/KEY: VARIANT
65 <222> LOCATION: 168
66 <223> OTHER INFORMATION: Xaa = Asp or Asn
W--> 68 <221> VARIANT
69 <222> LOCATION: 195, 196
70 <223> OTHER INFORMATION: Xaa = Ser or Ala
W--> 72 <221> VARIANT
73 <222> LOCATION: 205, 206
74 <223> OTHER INFORMATION: Xaa = Asp or Ala
W--> 76 <221> VARIANT
77 <222> LOCATION: 265, 266
78 <223> OTHER INFORMATION: Xaa = Ser or Thr
W--> 80 <221> VARIANT
81 <222> LOCATION: 358
82 <223> OTHER INFORMATION: Xaa = Glu or Gln
W--> 84 <400> 2
85 Met Arg Gly Lys Lys Val Trp Ile Ser Leu Leu Phe Ala Leu Ala Leu
86 1 5 10 15
87 Ile Phe Thr Met Ala Phe Gly Ser Thr Ser Ser Ala Gly Ala Ala Gly
88 20 25 30
89 Lys Ser Asn Gly Glu Lys Lys Tyr Ile Val Gly Phe Lys Gln Thr Met
90 35 40 45
91 Ser Thr Met Ser Ala Ala Lys Lys Lys Asp Val Ile Ser Glu Lys Gly
92 50 55 60
93 Gly Lys Val Gln Lys Gln Phe Lys Tyr Val Asp Ala Ala Ser Ala Thr
94 65 70 75 80
95 Leu Asn Glu Lys Ala Val Lys Glu Leu Lys Lys Asp Pro Ser Val Ala
96 85 90 95
97 Tyr Val Glu Glu Asp His Val Ala His Ala Tyr Ala Gln Ser Val Pro
98 100 105 110
99 Tyr Gly Val Ser Gln Ile Lys Ala Pro Ala Leu His Ser Gln Gly Tyr
100 115 120 125
101 Thr Gly Ser Asn Val Lys Val Ala Val Ile Asp Ser Gly Ile Asp Ser
102 130 135 140
103 Ser His Pro Asp Leu Lys Val Ala Gly Gly Ala Ser Met Val Pro Ser
104 145 150 155 160
W--> 105 Glu Thr Xaa Xaa Phe Gln Asp Xaa Asn Ser His Gly Thr His Val Ala
106 165 170 175
107 Gly Thr Val Ala Ala Leu Asn Asn Ser Ile Gly Val Leu Gly Val Ala
108 180 185 190
109 Pro Ser Xaa Xaa Leu Tyr Ala Val Lys Val Leu Gly Xaa Xaa Gly Ser
110 195 200 205
111 Gly Gln Tyr Ser Trp Ile Ile Asn Gly Ile Glu Trp Ala Ile Ala Asn
112 210 215 220

```

## RAW SEQUENCE LISTING

DATE: 09/29/2006

PATENT APPLICATION: US/10/500,936

TIME: 11:36:44

Input Set : A:\717\_2\_PCTseq.txt

Output Set: N:\CRF4\09292006\J500936.raw

```

113 Asn Met Asp Val Ile Asn Met Ser Leu Gly Gly Pro Ser Gly Ser Ala
114 225                230                235                240
115 Ala Leu Lys Ala Ala Val Asp Lys Ala Val Ala Ser Gly Val Val Val
116                245                250                255
117 Val Ala Ala Ala Gly Asn Glu Gly Xaa Xaa Gly Ser Ser Ser Thr Val
118                260                265                270
119 Gly Tyr Pro Gly Lys Tyr Pro Ser Val Ile Ala Val Gly Ala Val Asp
120                275                280                285
121 Ser Ser Asn Gln Arg Ala Ser Phe Ser Ser Val Gly Pro Glu Leu Asp
122                290                295                300
123 Val Met Ala Pro Gly Val Ser Ile Gln Ser Thr Leu Pro Gly Asn Lys
124 305                310                315                320
125 Tyr Gly Ala Tyr Asn Gly Thr Ser Met Ala Ser Pro His Val Ala Gly
126                325                330                335
127 Ala Ala Ala Leu Ile Leu Ser Lys His Pro Asn Trp Thr Asn Thr Gln
128                340                345                350
129 Val Arg Ser Ser Leu Xaa Asn Thr Thr Lys Leu Gly Asp Ser Phe
130                355                360                365
131 Tyr Tyr Gly Lys Gly Leu Ile Asn Val Gln Ala Ala Ala Gln
132                370                375                380
134 <210> SEQ ID NO: 3
135 <211> LENGTH: 275
136 <212> TYPE: PRT
137 <213> ORGANISM: Bacillus amyloliquefaciens
139 <400> SEQUENCE: 3
140 Ala Gln Ser Val Pro Tyr Gly Val Ser Gln Ile Lys Ala Pro Ala Leu
141 1                5                10                15
142 His Ser Gln Gly Tyr Thr Gly Ser Asn Val Lys Val Ala Val Ile Asp
143                20                25                30
144 Ser Gly Ile Asp Ser Ser His Pro Asp Leu Lys Val Ala Gly Gly Ala
145                35                40                45
146 Ser Met Val Pro Ser Glu Thr Asn Pro Phe Gln Asp Asn Asn Ser His
147 50                55                60
148 Gly Thr His Val Ala Gly Thr Val Ala Ala Leu Asn Asn Ser Ile Gly
149 65                70                75                80
150 Val Leu Gly Val Ala Pro Ser Ala Ser Leu Tyr Ala Val Lys Val Leu
151                85                90                95
152 Gly Ala Asp Gly Ser Gly Gln Tyr Ser Trp Ile Ile Asn Gly Ile Glu
153                100               105               110
154 Trp Ala Ile Ala Asn Asn Met Asp Val Ile Asn Met Ser Leu Gly Gly
155                115               120               125
156 Pro Ser Gly Ser Ala Ala Leu Lys Ala Ala Val Asp Lys Ala Val Ala
157 130               135               140
158 Ser Gly Val Val Val Val Ala Ala Ala Gly Asn Glu Gly Thr Ser Gly
159 145               150               155               160
160 Ser Ser Ser Thr Val Gly Tyr Pro Gly Lys Tyr Pro Ser Val Ile Ala
161                165               170               175
162 Val Gly Ala Val Asp Ser Ser Asn Gln Arg Ala Ser Phe Ser Ser Val
163                180               185               190

```

## RAW SEQUENCE LISTING

DATE: 09/29/2006

PATENT APPLICATION: US/10/500,936

TIME: 11:36:44

Input Set : A:\717\_2\_PCTseq.txt

Output Set: N:\CRF4\09292006\J500936.raw

```

164 Gly Pro Glu Leu Asp Val Met Ala Pro Gly Val Ser Ile Gln Ser Thr
165           195                200                205
166 Leu Pro Gly Asn Lys Tyr Gly Ala Tyr Asn Gly Thr Ser Met Ala Ser
167       210                215                220
168 Pro His Val Ala Gly Ala Ala Ala Leu Ile Leu Ser Lys His Pro Asn
169 225                230                235                240
170 Trp Thr Asn Thr Gln Val Arg Ser Ser Leu Glu Asn Thr Thr Thr Lys
171                245                250                255
172 Leu Gly Asp Ser Phe Tyr Tyr Gly Lys Gly Leu Ile Asn Val Gln Ala
173       260                265                270
174 Ala Ala Gln
175       275
177 <210> SEQ ID NO: 4
178 <211> LENGTH: 275
179 <212> TYPE: PRT
180 <213> ORGANISM: Bacillus subtilis
182 <400> SEQUENCE: 4
183 Ala Gln Ser Val Pro Tyr Gly Ile Ser Gln Ile Lys Ala Pro Ala Leu
184 1           5           10           15
185 His Ser Gln Gly Tyr Thr Gly Ser Asn Val Lys Val Ala Val Ile Asp
186       20           25           30
187 Ser Gly Ile Asp Ser Ser His Pro Asp Leu Asn Val Arg Gly Gly Ala
188       35           40           45
189 Ser Phe Val Pro Ser Glu Thr Asn Pro Tyr Gln Asp Gly Ser Ser His
190       50           55           60
191 Gly Thr His Val Ala Gly Thr Ile Ala Ala Leu Asn Asn Ser Ile Gly
192 65           70           75           80
193 Val Leu Gly Val Ser Pro Ser Ala Ser Leu Tyr Ala Val Lys Val Leu
194       85           90           95
195 Asp Ser Thr Gly Ser Gly Gln Tyr Ser Trp Ile Ile Asn Gly Ile Glu
196       100          105          110
197 Trp Ala Ile Ser Asn Asn Met Asp Val Ile Asn Met Ser Leu Gly Gly
198       115          120          125
199 Pro Thr Gly Ser Thr Ala Leu Lys Thr Val Val Asp Lys Ala Val Ser
200       130          135          140
201 Ser Gly Ile Val Val Ala Ala Ala Gly Asn Glu Gly Ser Ser Gly
202 145          150          155          160
203 Ser Thr Ser Thr Val Gly Tyr Pro Ala Lys Tyr Pro Ser Thr Ile Ala
204       165          170          175
205 Val Gly Ala Val Asn Ser Ser Asn Gln Arg Ala Ser Phe Ser Ser Ala
206       180          185          190
207 Gly Ser Glu Leu Asp Val Met Ala Pro Gly Val Ser Ile Gln Ser Thr
208       195          200          205
209 Leu Pro Gly Gly Thr Tyr Gly Ala Tyr Asn Gly Thr Ser Met Ala Thr
210       210          215          220
211 Pro His Val Ala Gly Ala Ala Ala Leu Ile Leu Ser Lys His Pro Thr
212 225          230          235          240
213 Trp Thr Asn Ala Gln Val Arg Asp Arg Leu Glu Ser Thr Ala Thr Tyr
214       245          250          255

```

## RAW SEQUENCE LISTING

DATE: 09/29/2006

PATENT APPLICATION: US/10/500,936

TIME: 11:36:44

Input Set : A:\717\_2\_PCTseq.txt

Output Set: N:\CRF4\09292006\J500936.raw

```

215 Leu Gly Asn Ser Phe Tyr Tyr Gly Lys Gly Leu Ile Asn Val Gln Ala
216           260           265           270
217 Ala Ala Gln
218           275
220 <210> SEQ ID NO: 5
221 <211> LENGTH: 274
222 <212> TYPE: PRT
223 <213> ORGANISM: Bacillus licheniformis
225 <400> SEQUENCE: 5
226 Ala Gln Thr Val Pro Tyr Gly Ile Pro Leu Ile Lys Ala Asp Lys Val
227 1           5           10           15
228 Gln Ala Gln Gly Phe Lys Gly Ala Asn Val Lys Val Ala Val Leu Asp
229           20           25           30
230 Thr Gly Ile Gln Ala Ser His Pro Asp Leu Asn Val Val Gly Gly Ala
231           35           40           45
232 Ser Phe Val Ala Gly Glu Ala Tyr Asn Thr Asp Gly Asn Gly His Gly
233           50           55           60
234 Thr His Val Ala Gly Thr Val Ala Ala Leu Asp Asn Thr Thr Gly Val
235 65           70           75           80
236 Leu Gly Val Ala Pro Ser Val Ser Leu Tyr Ala Val Lys Val Leu Asn
237           85           90           95
238 Ser Ser Gly Ser Gly Ser Tyr Ser Gly Ile Val Ser Gly Ile Glu Trp
239           100          105          110
240 Ala Thr Thr Asn Gly Met Asp Val Ile Asn Met Ser Leu Gly Gly Ala
241           115          120          125
242 Ser Gly Ser Thr Ala Met Lys Gln Ala Val Asp Asn Ala Tyr Ala Arg
243           130          135          140
244 Gly Val Val Val Val Ala Ala Ala Gly Asn Ser Gly Asn Ser Gly Ser
245 145          150          155          160
246 Thr Asn Thr Ile Gly Tyr Pro Ala Lys Tyr Asp Ser Val Ile Ala Val
247           165          170          175
248 Gly Ala Val Asp Ser Asn Ser Asn Arg Ala Ser Phe Ser Ser Val Gly
249           180          185          190
250 Ala Glu Leu Glu Val Met Ala Pro Gly Ala Gly Val Tyr Ser Thr Tyr
251           195          200          205
252 Pro Thr Asn Thr Tyr Ala Thr Leu Asn Gly Thr Ser Met Ala Ser Pro
253           210          215          220
254 His Val Ala Gly Ala Ala Ala Leu Ile Leu Ser Lys His Pro Asn Leu
255 225          230          235          240
256 Ser Ala Ser Gln Val Arg Asn Arg Leu Ser Ser Thr Ala Thr Tyr Leu
257           245          250          255
258 Gly Ser Ser Phe Tyr Tyr Gly Lys Gly Leu Ile Asn Val Glu Ala Ala
259           260          265          270
260 Ala Gln
263 <210> SEQ ID NO: 6
264 <211> LENGTH: 269
265 <212> TYPE: PRT
266 <213> ORGANISM: Bacillus lentus
268 <400> SEQUENCE: 6

```

RAW SEQUENCE LISTING ERROR SUMMARY  
PATENT APPLICATION: US/10/500,936

DATE: 09/29/2006  
TIME: 11:36:45

Input Set : A:\717\_2\_PCTseq.txt  
Output Set: N:\CRF4\09292006\J500936.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:2; Xaa Pos. 163,164,168,195,196,205,206,265,266,358  
Seq#:9; Xaa Pos. 6,8,9,10,11,12,16,18,19,21,25,31,33,35,36,37,38,43,44,45  
Seq#:9; Xaa Pos. 50,53,54,55,56,57,58,61,63,72,89,99,103,104,106,108,109  
Seq#:9; Xaa Pos. 111,115,117,118,119,120,122,124,128,131,133,134,135,136  
Seq#:9; Xaa Pos. 137,139,140,141,142,143,144,145,148,153,156,158,159,160  
Seq#:9; Xaa Pos. 161,162,163,164,165,166,169,170,172,173,174,175,180,182  
Seq#:9; Xaa Pos. 183,185,186,191,192,194,195,198,199,204,205,209,212,213  
Seq#:9; Xaa Pos. 215,216,217,224,234,235,236,238,239,240,242,243,244,246  
Seq#:9; Xaa Pos. 248,249,251,254,255,256,259,260,261,262,265,268,270,271  
Seq#:9; Xaa Pos. 274,275

## VERIFICATION SUMMARY

DATE: 09/29/2006

PATENT APPLICATION: US/10/500,936

TIME: 11:36:45

Input Set : A:\717\_2\_PCTseq.txt

Output Set: N:\CRF4\09292006\J500936.raw

L:12 M:270 C: Current Application Number differs, Replaced Current Application Number  
L:13 M:271 C: Current Filing Date differs, Replaced Current Filing Date  
L:68 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:2  
L:72 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:2  
L:76 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:2  
L:80 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:2  
L:84 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:2  
L:105 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:2 after pos.:160  
M:341 Repeated in SeqNo=2  
L:334 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!  
L:338 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:9  
L:339 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:9 after pos.:0  
M:341 Repeated in SeqNo=9